ISIS: A two-scale framework to sparse inference

Jianqing Fan

Princeton University

http://www.princeton.edu/∼jqfan

March 26, 2010

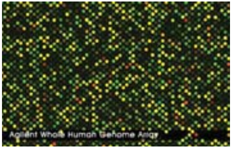
Outline

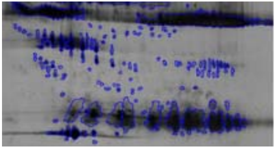
1. Introduction
2. Challenge of Dimensionality
3. The ISIS Method
4. Sparse Survival analysis
5. Numerical Examples

Examples: Biological Sciences and Engineering

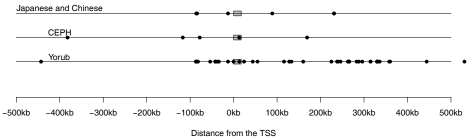
High-dim variable selection characterizes many contemporary statistical problems.

- Bioinformatic: disease classification / predicting clinical outcomes using microarray, proteomics, fMRI data;





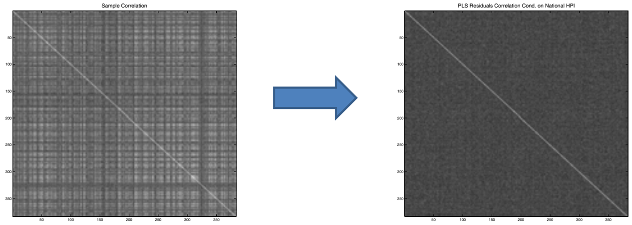
- Association between phenotypes and SNPs or eQTL



- Document or text classification: E-mail spam.

Example: Economics, Finance, Marketing

- HPIs / drug sales collected in many regions
- Local correlation makes dimensionality growths quickly.

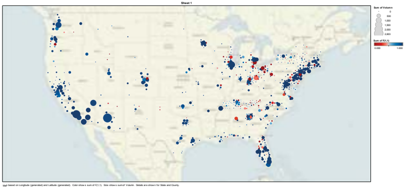


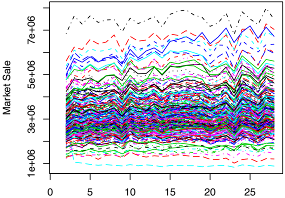
- 1000 neighborhoods requires 1 m parameters.
- Managing 2K stocks involves 2m elements in covariance.

Example: Spatial temporal data

■Meteorology & Earth Sciences & Ecology

- Temperatures and other attributes (precipitation, population size) are collected over time and over many regions.





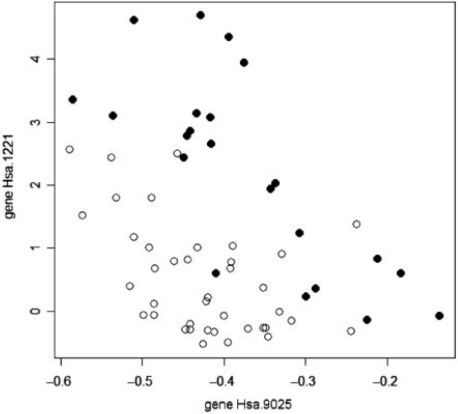
- Large panel data over a short time horizon poses more challenges.

Growth of Dimensionality

- Dimen. grows rapidly w/ interactions: 5000 → 12.5m.

Synergy of Two Genes: colon cancer in Hanczar et al (2007).

e.g., $Y = I(X_1 + X_2 > 3)$ **and** $Y \perp X_1$.

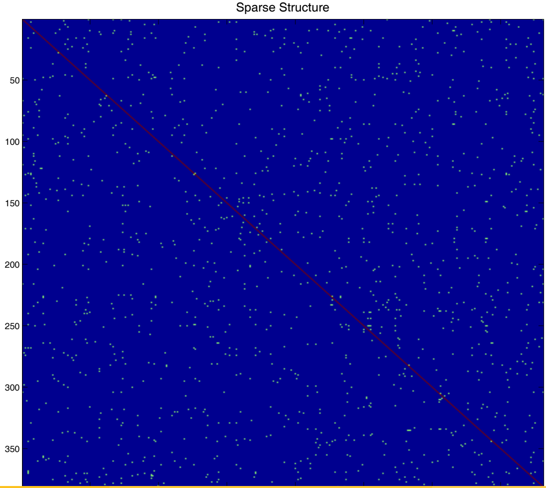


white – patients; black – normal

Essential Assumption: Sparsity

Dimensionality: $\log p = O(n^a)$ (**NP**-dimensionality)

Intrinsic dimensionality: $s \ll n$. (Sparsity)

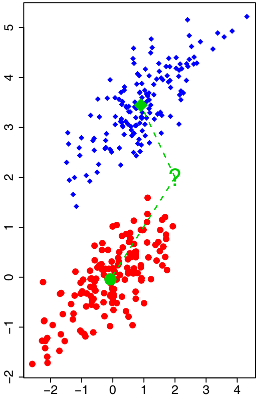


Impact of High-dimensionality

Impact of Dimensionality

Regression:

- **Not** directly implementable if $p > n$.
- Prediction error is $(1 + \frac{p}{n})\sigma^2$, if $p \leq n$.



Classification: No implementation problems, but **error rates**

—depend on $C_p^2/\sqrt{p}$ (*Fan & Fan 08*), C_p is **distance**.

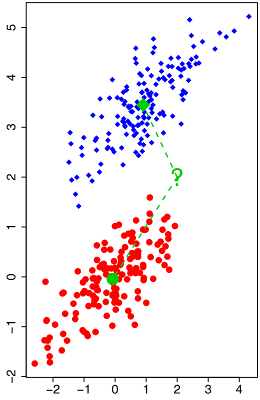
—**perfectly classifiable** if $C_p^2/\sqrt{p} \longrightarrow \infty$ (*Hall, Pittelkow & Ghosh,08*).

Impact of Dimensionality

Regression:

■**Not** directly implementable if $p > n$.

■Prediction error is $(1 + \frac{\mathbf{p}}{\mathbf{n}})\sigma^2$, if $p \leq n$.



Classification: No implementation problems, but **error rates**

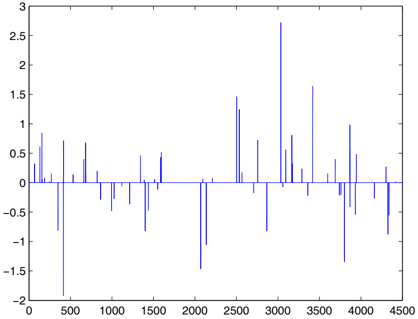
—depend on $C_p^2/\sqrt{p}$ (*Fan & Fan 08*), C_p is **distance**.

—**perfectly classifiable** if $C_p^2/\sqrt{p} \to \infty$ (*Hall, Pittelkow & Ghosh,08*).

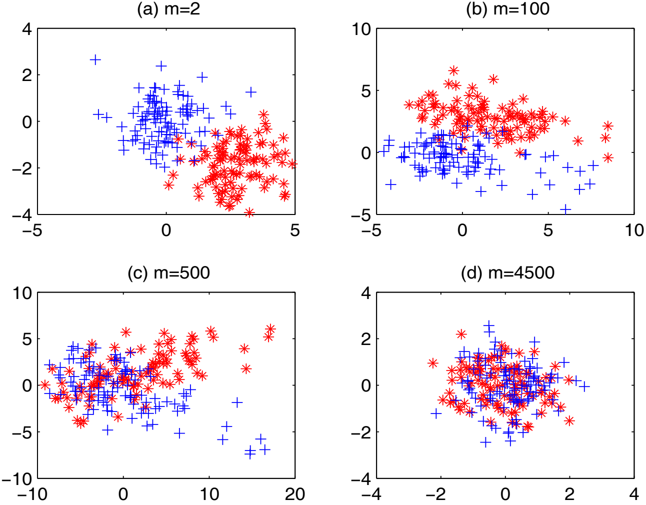
An illustration

■**dimensionality**: $p = 4500, \qquad n = 200$

■**Signals**: $\mu_1 = 0.98\delta_0 + 0.02\text{DE}, \qquad \mu_2 = 0$



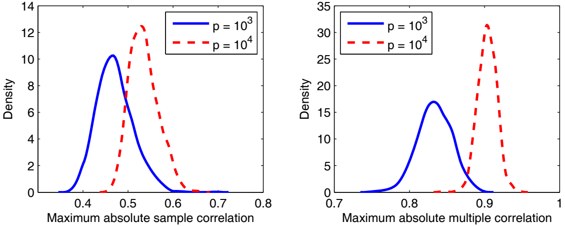
Impact of Dimensionality on classification



■ Classification power depends on $\sum_{i=1}^{m} \alpha_i^2 / \sqrt{m}$.

Spurious Relationships

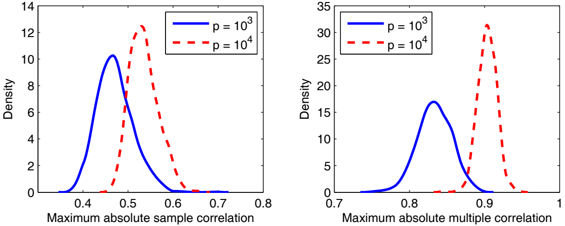
An experiment: Generate $n = 50$ $Z_1, \cdots, Z_p \sim_{i.i.d.} N(0,1)$; ■compute $r = \max_{j \geq 2} \text{corr}(Z_1, Z_j)$.



■compute maximum multiple correlation: $R = \max_{|S|=5} \text{corr}(Z_1, \mathbf{Z}_S)$.

Spurious Relationships

An experiment: Generate $n = 50$ $Z_1, \cdots, Z_p \sim_{i.i.d.} N(0, 1)$; ■compute $r = \max_{j \geq 2} \text{corr}(Z_1, Z_j)$.



■compute maximum multiple correlation: $R = \max_{|S|=5} \text{corr}(Z_1, \mathbf{Z}_S)$.

False Outcomes

Scientific implication: If Z_1 is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

False statistical inferences: If $Y = Z_1$ and fit

$$Y = \mathbf{Z}_{\hat{S}}^T \beta + \varepsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - 0.9^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than 1/2

★ more variables to be called "statistically significant".

False Outcomes

Scientific implication: If Z_1 is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

False statistical inferences: If $Y = Z_1$ and fit

$$Y = \mathbf{Z}_{\hat{S}}^T \beta + \varepsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - \mathbf{0.9}^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than 1/2

★ more variables to be called "statistically significant".

False Outcomes

Scientific implication: If Z_1 is responsible for breast cancer, but we can also discover other 5 genes, **indep of outcome**!

False statistical inferences: If $Y = Z_1$ and fit

$$Y = \mathbf{Z}_{\hat{S}}^T \beta + \varepsilon,$$

the residual variance

$$\hat{\sigma}^2 = \frac{RSS}{n - |\hat{S}|} \approx (1 - \mathbf{0.9}^2) \times \frac{49}{45} \approx 0.2,$$

★ $\hat{\sigma}$ is deflation by a factor of more than 1/2

★ more variables to be called "statistically significant".

Curse of Ultrahigh Dimensionality

- Computational cost
- Estimation accuracy.
- Stability
- Noise Accumulation

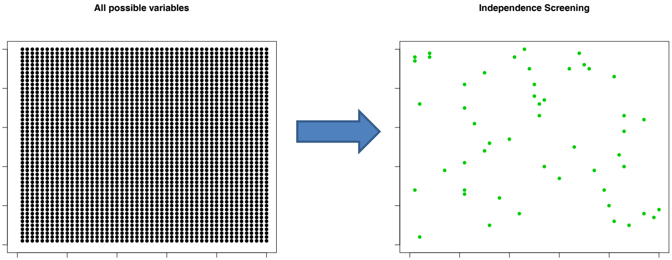
Key Idea: **Large-scale** screening + **moderate-scale** searching.

The ISIS Method

a two-scale framework

Hydrogen Atom: Large scale-screening

Indep learning: Feature ranking by **Marginal** correlation (*Fan & Lv, 08*) or generalized correlation (*Hall & Miller, 09*);



Classification: Feature ranking by two-sample t-tests or other tests (Tibshirani, et al, 03; Fan and Fan, 2008).

Extensions & Questions

Other methods: ★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MMLE** (*Fan and Song, 09*); ★**MPLE** (*Zhao & Li, 09*); ★**Nonparametric learning** (*Fan, Feng, Song, 09*) ★**Data-tilting**; (*Hall, Titterington & Xue, 09*).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?

Extensions & Questions

Other methods: ★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MMLE** (*Fan and Song, 09*); ★**MPLE** (*Zhao & Li, 09*); ★**Nonparametric learning** (*Fan, Feng, Song, 09*) ★**Data-tilting**; (*Hall, Titterington & Xue, 09*).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?

Extensions & Questions

Other methods: ★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MMLE** (*Fan and Song, 09*); ★**MPLE** (*Zhao & Li, 09*); ★**Nonparametric learning** (*Fan, Feng, Song, 09*) ★**Data-tilting**; (*Hall, Titterington & Xue, 09*).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?

Extensions & Questions

Other methods: ★**Marginal LR** (*Fan, Samworth & Wu, 09*); ★**MMLE** (*Fan and Song, 09*); ★**MPLE** (*Zhao & Li, 09*); ★**Nonparametric learning** (*Fan, Feng, Song, 09*) ★**Data-tilting**; (*Hall, Titterington & Xue, 09*).

1. Can we have model selection consistency?
2. Can we have sure screening property? In what capacity?
3. How to choose a thresholding parameter?
4. How to reduce FDR?
5. What are the possible drawbacks?

Potential Drawbacks

- **False Negative**: What if X_4 marginally uncorrelated with Y, but jointly correlated with Y?

$$Y = X_1 + X_2 + X_3 + \beta_4 \mathbf{X_4} + \varepsilon \quad \text{s.t.} \quad \mathrm{cov}(Y, X_4) = 0.$$

- **False Positive**: What if $X_2, \cdots, X_{99}$ highly correlated with an important X_1, but weakly correlated with Y conditionally?

$$Y = X_1 + \mathbf{0.2} X_{100} + \varepsilon$$

Potential Drawbacks

- **False Negative**: What if X_4 marginally uncorrelated with Y, but jointly correlated with Y?

$$Y = X_1 + X_2 + X_3 + \beta_4 \mathbf{X_4} + \varepsilon \quad \text{s.t.} \quad \mathrm{cov}(Y, X_4) = 0.$$

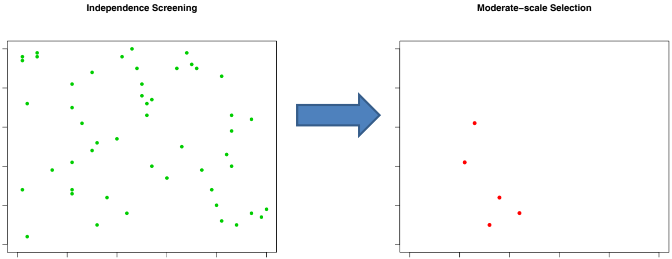
- **False Positive**: What if $X_2, \cdots, X_{99}$ highly correlated with an important X_1, but weakly correlated with Y conditionally?

$$Y = X_1 + \mathbf{0.2} X_{100} + \varepsilon$$

Oxygen Atom: Penalized likelihood estimation

$$Q(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, \mathbf{x}_{i,d}^T \beta) + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

■ Simultaneously estimate coefs and choose variables.

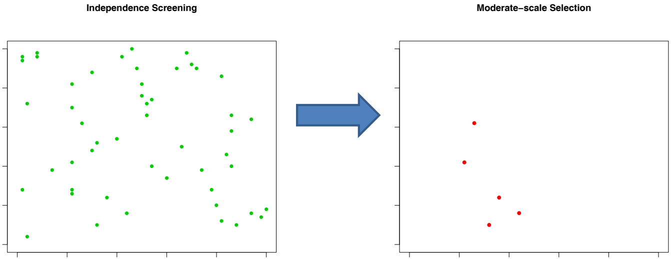


- How high dimensionality can such methods handle?
- What is the role of penalty functions?
- Does it possess an oracle property? How to compute?

Oxygen Atom: Penalized likelihood estimation

$$Q(\beta) = n^{-1}\sum_{i=1}^{n} L(Y_i, \mathbf{x}_{i,d}^T\beta) + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

■ Simultaneously estimate coefs and choose variables.



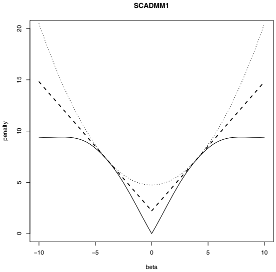
- How high dimensionality can such methods handle?
- What is the role of penalty functions?
- Does it possess an oracle property? How to compute?

Penalized likelihood estimation

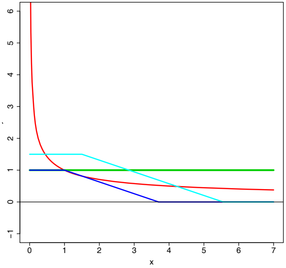
Penalty: Popular choice L_1. Preferred choice: SCAD or MCP.

■ Better bias property and model selection consistency.

$$Q^{\text{app}}(\beta) = n^{-1}\sum_{i=1}^{n} L(Y_i, \mathbf{x}_{i,d}^T\beta) + \sum_{j=1}^{d} w_j|\beta|_j, \qquad w_j = p'_\lambda(|\beta_j^{(k)}|)$$



,



Oracle property: can handle NP-dimensionality.

Computation: **LQA** (*Fan & Li, 01*); **LLA** (*Zou & Li, 08*);

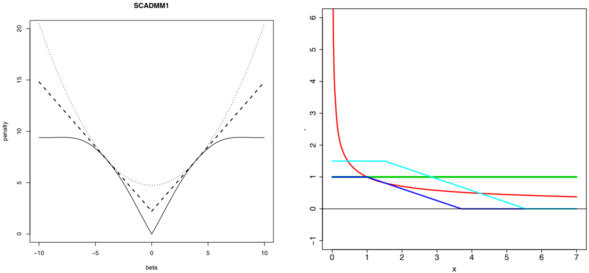
PLUS (*Zhang, 09*); **Coordinate optimization** (*Fu & Jiang, 99*).

Penalized likelihood estimation

Penalty: Popular choice L_1. Preferred choice: SCAD or MCP.

■Better bias property and model selection consistency.

$$Q^{\text{app}}(\beta) = n^{-1} \sum_{i=1}^{n} L(Y_i, \mathbf{x}_{i,d}^T \beta) + \sum_{j=1}^{d} w_j |\beta|_j, \qquad w_j = p'_\lambda(|\beta_j^{(k)}|)$$



,

Oracle property: can handle NP-dimensionality.

Computation: **LQA** (*Fan & Li, 01*); **LLA** (*Zou & Li, 08*);

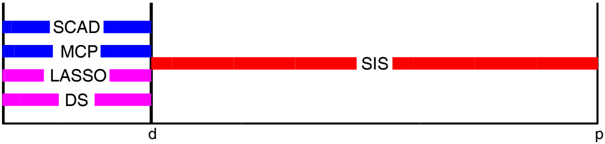
PLUS (*Zhang, 09*); **Coordinate optimization** (*Fu & Jiang, 99*).

Carbon Atom

Iterative application of

large-scale **screening** and

moderate-scale **selection**.



■ISIS (*(Fan & Lv, 08; Fan, Samworth & Wu, 09)*), **available in R**.

Iterative conditional feature selection

1. ■**(Large-scale screening)**: Apply SIS to pick a set $\mathcal{A}_1$;
   ■**(Moderate-scale selection)**: Employ a penalized likelihood to select a subset $\mathcal{M}_1$ of these indices.

2. **(Large-scale screening)**: Rank features according to the additional (**conditional**) contribution:

$$L_j^{(2)} = \min_{\beta_0, \beta_{\mathcal{M}_1}, \beta_j} n^{-1} \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_{\mathbf{i}, \mathcal{M}_1}^{\mathbf{T}} \beta_{\mathcal{M}_1} + X_{ij}\beta_j),$$

   resulting in $\mathcal{A}_2$.
   —Improvement of **residual** approach (FL 08): $\beta_{\mathcal{M}_1} = \hat{\beta}_{\mathcal{M}_1}$.

Iterative conditional feature selection

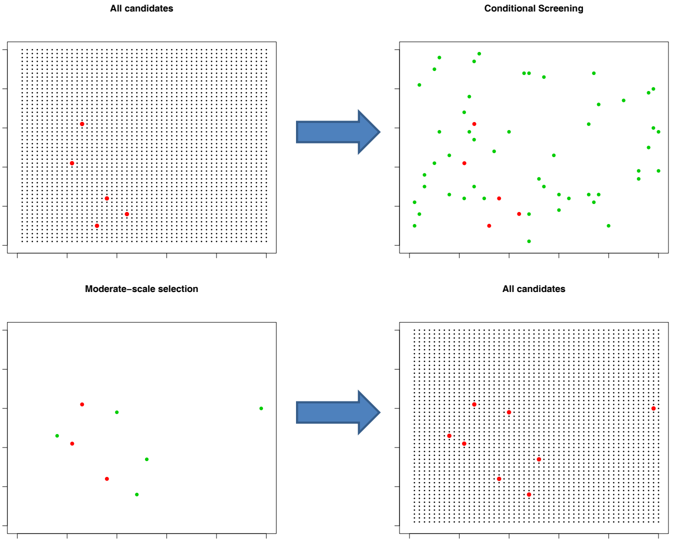
1. **(Large-scale screening)**: Apply SIS to pick a set $\mathcal{A}_1$;
   (Moderate-scale selection): Employ a penalized likelihood to select a subset $\mathcal{M}_1$ of these indices.

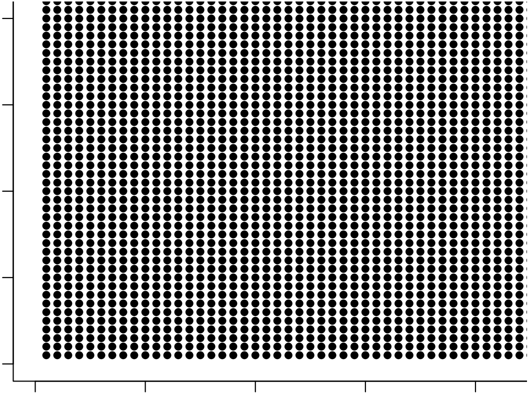
2. **(Large-scale screening)**: Rank features according to the additional (**conditional**) contribution:

$$L_j^{(2)} = \min_{\beta_0, \beta_{\mathcal{M}_1}, \beta_j} n^{-1} \sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_{\mathbf{i},\mathcal{M}_1}^{\mathbf{T}} \beta_{\mathcal{M}_1} + X_{ij}\beta_j),$$

   resulting in $\mathcal{A}_2$.
   —Improvement of **residual** approach (FL 08): $\beta_{\mathcal{M}_1} = \hat{\beta}_{\mathcal{M}_1}$.

Illustration of ISIS





Iterative feature selection (II)

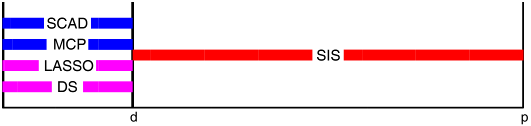
3. **(Moderate-scale selection)**: Minimize wrt $\beta_{\mathcal{M}_1}, \beta_{\mathcal{A}_2}$

$$\sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_{i,\mathcal{M}_1}^T \beta_{\mathcal{M}_1} + \mathbf{x}_{i,\mathcal{A}_2}^T \beta_{\mathcal{A}_2}) + \sum_{j \in \mathcal{M}_1 \cup \mathcal{A}_2} p_\lambda(|\beta_j|),$$

resulting in $\mathcal{M}_2$

—**Allow deletion**, improvement over ISIS (Fan and Lv, 08).

4. Repeat Steps 1–3 until $|\mathcal{M}_\ell| = d$ (prescribed) or $\mathcal{M}_\ell = \mathcal{M}_{\ell-1}$.



Iterative feature selection (II)

3. **(Moderate-scale selection)**: Minimize wrt $\beta_{\mathcal{M}_1}$, $\beta_{\mathcal{A}_2}$

$$\sum_{i=1}^{n} L(Y_i, \beta_0 + \mathbf{x}_{i,\mathcal{M}_1}^T \beta_{\mathcal{M}_1} + \mathbf{x}_{i,\mathcal{A}_2}^T \beta_{\mathcal{A}_2}) + \sum_{j \in \mathcal{M}_1 \cup \mathcal{A}_2} p_\lambda(|\beta_j|),$$

resulting in $\mathcal{M}_2$

—Allow deletion, improvement over ISIS (Fan and Lv, 08).

4. Repeat Steps 1–3 until $|\mathcal{M}_\ell| = d$ (prescribed) or $\mathcal{M}_\ell = \mathcal{M}_{\ell-1}$.

Applicability of ISIS idea

The idea of ISIS is widely applicable. It can be applied to

- Classification (*Fan, Samworth, & Wu, 09*).
- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).
- Nonparametric learning (*Fan, Feng, & Song, 09*).
- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)

Applicability of ISIS idea

The idea of ISIS is widely applicable. It can be applied to

- Classification (*Fan, Samworth, & Wu, 09*).
- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).
- Nonparametric learning (*Fan, Feng, & Song, 09*).
- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)

Applicability of ISIS idea

The idea of ISIS is widely applicable. It can be applied to

- Classification (*Fan, Samworth, & Wu, 09*).
- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).
- Nonparametric learning (*Fan, Feng, & Song, 09*).
- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)

Applicability of ISIS idea

The idea of ISIS is widely applicable. It can be applied to

- Classification (*Fan, Samworth, & Wu, 09*).
- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).
- Nonparametric learning (*Fan, Feng, & Song, 09*).
- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)

Applicability of ISIS idea

The idea of ISIS is widely applicable. It can be applied to

- Classification (*Fan, Samworth, & Wu, 09*).
- **Survival analysis** (*Zhao & Li, 09; Fan, Feng, & Wu, 09*).
- Nonparametric learning (*Fan, Feng, & Song, 09*).
- Robust and quantile regression (*Bradic, Fan, & Wang, 09*)

Sparse Survival Analysis

Hazard regression model

Notation:

- Survival time: T. Censoring time: C.
- Observation time: $Y = \min(T, C)$
- Censoring indicator $\delta = I(T \leq C)$.
- Covariates: $\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$.

Cox's proportional hazards model (Cox, 1972 & 1975):

$$h(t|\mathbf{x}) = h_0(t)\Psi(\mathbf{x}) = \mathbf{h_0(t)}\exp(\mathbf{x}^T\beta). \qquad \blacksquare\ \beta \text{ sparse.}$$

Hazard regression model

Notation:

- Survival time: T. Censoring time: C.
- Observation time: $Y = \min(T, C)$
- Censoring indicator $\delta = I(T \le C)$.
- Covariates: $\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$.

Cox's proportional hazards model (Cox, 1972 & 1975):

$$h(t|\mathbf{x}) = h_0(t)\Psi(\mathbf{x}) = \mathbf{h_0(t)}\exp(\mathbf{x}^{\mathbf{T}}\beta).$$

■ β sparse.

Marginal partial likelihood

Marginal utility: Screening according to $|\hat{\beta}_j^M|$ or

$$u_j = \max_{\beta_j} \left(\sum_{i=1}^n \delta_i x_{ij}\beta_j - \sum_{i=1}^n \delta_i \log\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{mj}\beta_j)\} \right).$$

Thresholding parameter: Zhao and Li (09) proposed using upper α **(Control of FDR)** quantile of marginal utilities for decoupled response and covariate (PSIS).

Theoretical properties: sure independence screening property with false positive rate controlled by α. (*Zhao and Li, 09*)

Marginal partial likelihood

Marginal utility: Screening according to $|\hat{\beta}_j^M|$ or

$$u_j = \max_{\beta_j} \left(\sum_{i=1}^n \delta_i x_{ij} \beta_j - \sum_{i=1}^n \delta_i \log\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{mj}\beta_j) \} \right).$$

Thresholding parameter: Zhao and Li (09) proposed using upper α **(Control of FDR)** quantile of marginal utilities for decoupled response and covariate (PSIS).

Theoretical properties: sure independence screening property with false positive rate controlled by α. (*Zhao and Li, 09*)

Marginal partial likelihood

Marginal utility: Screening according to $|\hat{\beta}_j^M|$ or

$$u_j = \max_{\beta_j} \left(\sum_{i=1}^n \delta_i x_{ij}\beta_j - \sum_{i=1}^n \delta_i \log\{ \sum_{m \in \mathcal{R}(y_i)} \exp(x_{mj}\beta_j) \} \right).$$

Thresholding parameter: Zhao and Li (09) proposed using upper α **(Control of FDR)** quantile of marginal utilities for decoupled response and covariate (PSIS).

Theoretical properties: sure independence screening property with false positive rate controlled by α. (*Zhao and Li, 09*)

Moderate Scale Selection

Among active variables I that pass the screening process, select them by **Penalized partial likelihood**:

$$\min_{\beta_I} \left(-\sum_{i=1}^{n} \delta_i \mathbf{x}_{I,i}^T \beta_I + \sum_{i=1}^{n} \delta_i \log\{ \sum_{m \in \mathcal{R}(y_i)} \exp(\mathbf{x}_{m,I}^T \beta_I) \} + \sum_{\mathbf{j} \in I} \mathbf{p}_\lambda(\beta_\mathbf{j}) \right),$$

Reduce the active set further to $\widehat{\mathcal{M}}_1$.

Conditional screening: **All** variables are candidates again. Rank them by conditional contribution given variables in $\widehat{\mathcal{M}}_1$.

Moderate Scale Selection

Among active variables I that pass the screening process, select them by **Penalized partial likelihood**:

$$\min_{\beta_I} \left(- \sum_{i=1}^{n} \delta_i \mathbf{x}_{I,i}^T \beta_I + \sum_{i=1}^{n} \delta_i \log\{ \sum_{m \in \mathcal{R}(y_i)} \exp(\mathbf{x}_{m,I}^T \beta_I) \} + \sum_{j \in I} \mathbf{p}_\lambda(\beta_j) \right),$$

Reduce the active set further to $\widehat{\mathcal{M}}_1$.

Conditional screening: **All** variables are candidates again. Rank them by conditional contribution given variables in $\widehat{\mathcal{M}}_1$.

Simulation Studies

Fan, Feng and Wu (2010)

Simulation settings

- $n = 300$ and $p = 400$, baseline hazard $h_0(t) = 0.1$.
- Designs: $X_1, \cdots, X_p$ are normal, equally correlated.
  - ★ Case 1: $\rho = 0$.
  - ★ Case 2: $\rho = 0.5$.
  - ★ Case 3: $\rho = 1/\sqrt{2}$.
  - ★ Case 4: case 3 + an indep variable.
- Easy to difficulty: case 1 to case 4.
- Hidden signature variable: Large impact but indep. of outcome.

Simulation settings

- $n = 300$ and $p = 400$, baseline hazard $h_0(t) = 0.1$.
- Designs: $X_1, \cdots, X_p$ are normal, equally correlated.
  - ★ Case 1: $\rho = 0$.
  - ★ Case 2: $\rho = 0.5$.
  - ★ Case 3: $\rho = 1/\sqrt{2}$.
  - ★ Case 4: case 3 $+$ an indep variable.
- Easy to difficulty: case 1 to case 4.
- Hidden signature variable: Large impact but indep. of outcome.

Simulation settings

- $n = 300$ and $p = 400$, baseline hazard $h_0(t) = 0.1$.
- Designs: $X_1, \cdots, X_p$ are normal, equally correlated.
  - ★ Case 1: $\rho = 0$.
  - ★ Case 2: $\rho = 0.5$.
  - ★ Case 3: $\rho = 1/\sqrt{2}$.
  - ★ Case 4: case 3 $+$ an indep variable.
- Easy to difficulty: case 1 to case 4.
- **Hidden signature variable**: Large impact but indep. of outcome.

Design of Simulation

Case 1 & 2: $\beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70$.

Case 3: $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -\mathbf{6\sqrt{2}}$.
Hidden siganature variable: X_4, indep $\mathbf{X}^T\beta^\star$ and T.

Case 4: $\beta_1 = \beta_2 = \beta_3 = 4$, $\beta_4 = -\mathbf{6\sqrt{2}}$ and $\beta_5 = \mathbf{4/3}$. **Weak variable**: X_5, marginally ranked below others.

★ **Challenging**: Case 4 contains a hidden signature variable (**false negative**) and a weak variable (**false positive**).

★ Censoring rate: about 30%

Design of Simulation

Case 1 & 2: $\beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70$.

Case 3: $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -\mathbf{6\sqrt{2}}$.
Hidden siganature variable: X_4, indep $\mathbf{X}^T\beta^\star$ and T.

Case 4: $\beta_1 = \beta_2 = \beta_3 = 4$, $\boldsymbol{\beta_4 = -6\sqrt{2}}$ and $\boldsymbol{\beta_5 = 4/3}$. **Weak variable**: X_5, marginally ranked below others.

★ **Challenging**: Case 4 contains a hidden signature variable (**false negative**) and a weak variable (**false positive**).

★ Censoring rate: about 30%

Design of Simulation

Case 1 & 2: $\beta_1 = -1.63, \beta_2 = 1.34, \beta_3 = -1.65, \beta_4 = 1.65, \beta_5 = -1.42, \beta_6 = 1.70$.

Case 3: $\beta_1 = \beta_2 = \beta_3 = 4$ and $\beta_4 = -\mathbf{6\sqrt{2}}$.
Hidden siganature variable: X_4, indep $\mathbf{X}^T\beta^\star$ and T.

Case 4: $\beta_1 = \beta_2 = \beta_3 = 4$, $\beta_4 = -\mathbf{6\sqrt{2}}$ and $\beta_5 = \mathbf{4/3}$. **Weak variable**: X_5, marginally ranked below others.

★ **Challenging**: Case 4 contains a hidden signature variable (**false negative**) and a weak variable (**false positive**).

★ Censoring rate: about 30%

Results for Cases 1 and 2 (100 simulations)

Case 1: independent covariates			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**0.79**	**0.57**	**4.23**
$\|\beta - \hat{\beta}\|_2^2$	0.13	0.09	0.98
Proportion	1	1	1
MMS	**7**	**6**	**68.5**
Case 2: Equi-correlated covariates with $\rho = 0.5$			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**2.2**	**0.64**	**4.38**
$\|\beta - \hat{\beta}\|_2^2$	1.74	0.11	0.98
Proportion	0.71	1	1
MMS	**7**	**6**	**57**

Results for Cases 1 and 2 (100 simulations)

Case 1: independent covariates			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**0.79**	**0.57**	**4.23**
$\|\beta - \hat{\beta}\|_2^2$	0.13	0.09	0.98
Proportion	1	1	1
MMS	**7**	**6**	**68.5**

Case 2: Equi-correlated covariates with $\rho = 0.5$			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**2.2**	**0.64**	**4.38**
$\|\beta - \hat{\beta}\|_2^2$	1.74	0.11	0.98
Proportion	0.71	1	1
MMS	**7**	**6**	**57**

Results for Cases 3 and 4

Case 3: A hidden signature variable			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**20.1**	**1.03**	**20.53**
$\|\beta - \hat{\beta}\|_2^2$	94.72	0.49	76.31
Proportion	0	1	0.06
MMS	**13**	**4**	**118.5**

Case 4: Two very hard variables to be selected.			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**20.87**	**1.15**	**21.04**
$\|\beta - \hat{\beta}\|_2^2$	96.46	0.51	77.03
Proportion	0	1	0.02
MMS	**13**	**5**	**118**

Results for Cases 3 and 4

Case 3: A hidden signature variable			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**20.1**	**1.03**	**20.53**
$\|\beta - \hat{\beta}\|_2^2$	94.72	0.49	76.31
Proportion	0	1	0.06
MMS	**13**	**4**	**118.5**
Case 4: Two very hard variables to be selected.			
	SIS	ISIS	LASSO
$\|\beta - \hat{\beta}\|_1$	**20.87**	**1.15**	**21.04**
$\|\beta - \hat{\beta}\|_2^2$	96.46	0.51	77.03
Proportion	0	1	0.02
MMS	**13**	**5**	**118**

Computation costs

■ Average running time (in seconds) for Van-ISIS and LASSO.

	Case 1	Case 2	Case 3	Case 4
Van-ISIS	379.29	213.44	402.94	231.68
LASSO	37730.82	26348.12	46847.00	28157.71

Cases 2 and 4 with $p = 1000$

Case 2 with $p = 1000$ and $n = 400$		
	SIS	ISIS
$\|\beta - \hat{\beta}\|_1$	**1.53**	**0.52**
$\|\beta - \hat{\beta}\|_2^2$	0.9	0.07
Proportion	**0.82**	**1**
MMS	8	6

Case 4 with $p = 1000$ and $n = 400$		
	SIS	ISIS
$\|\beta - \hat{\beta}\|_1$	**20.88**	**0.99**
$\|\beta - \hat{\beta}\|_2^2$	93.53	0.39
Proportion	**0**	**1**
MMS	16	5

Cases 2 and 4 with $p = 1000$

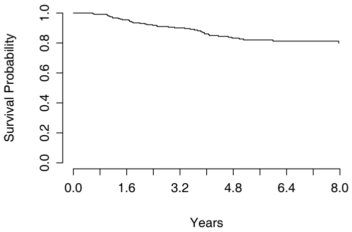
Case 2 with $p = 1000$ and $n = 400$		
	SIS	ISIS
$\|\beta - \hat{\beta}\|_1$	**1.53**	**0.52**
$\|\beta - \hat{\beta}\|_2^2$	0.9	0.07
Proportion	**0.82**	**1**
MMS	8	6
Case 4 with $p = 1000$ and $n = 400$		
	SIS	ISIS
$\|\beta - \hat{\beta}\|_1$	**20.88**	**0.99**
$\|\beta - \hat{\beta}\|_2^2$	93.53	0.39
Proportion	**0**	**1**
MMS	16	5

An Application

Fan, Feng and Wu (2010)

Neuroblastoma data set

- Purpose: study genes related to survival information for German Neuroblastoma Trials.
- $n = 251, p = 10,707$, Affymetrix arrays.
- The censoring rate is 205/246.
- Survival curve of 246 patients is given by

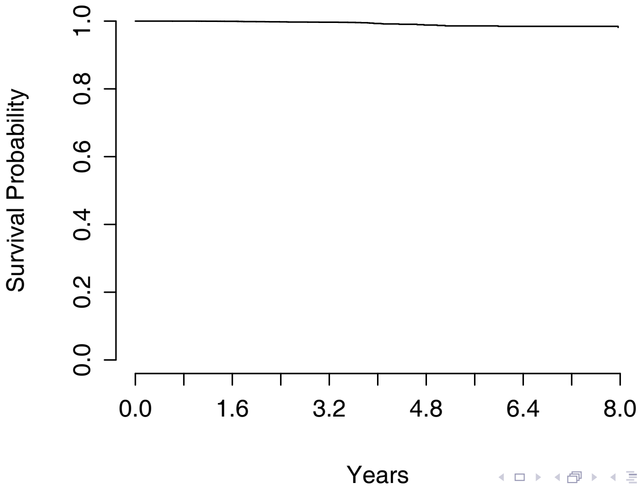


Selected Genes and its impact

Probe ID	estimated coefficient	standard error	p-value
A_23_P31816	0.864	0.203	2.1e-05
A_23_P31816	-0.940	0.314	2.8e-03
A_23_P31816	-0.815	1.704	6.3e-01
A_32_P424973	-1.957	0.396	7.8e-07
A_32_P159651	-1.295	0.185	2.6e-12
Hs61272.2	1.664	0.249	2.3e-11
Hs13208.1	-0.789	0.149	1.1e-07
Hs150167.1	1.708	1.687	3.1e-01

- Six genes are significant and not removable.
- Adding two random selected genes 20 times, **no pairs** have significant contributions

Estimated baseline survival function



Acknowledgement

Thank  *You*

In collaboration with

- ★ Yang Feng (*Princeton University; FFW, 2009*)
- ★ Jinchi Lv (*University of Southern California; Fan & Lv; 2008*)
- ★ Richard Samworth (*Cambridge University; FSW, 2009*).
- ★ Rui Song (*Colorado State University, Fan & Song, 2009*).
- ★ Yichao Wu (*North Carolina State University, FSW, 2009*).